

Weighted Ensemble Path Sampling for Multiple Reaction Channels

Bin W. Zhang,¹ David Jasnow,² and Daniel M. Zuckerman ^{*1}

¹*Department of Computational Biology, School of Medicine,
University of Pittsburgh, Pennsylvania 15213*

²*Department of Physics & Astronomy,
University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

(Dated: February 16, 2009)

Abstract

Finding and sampling multiple reaction channels for molecular transitions remains an important challenge in physical chemistry. Here we show that the weighted ensemble (WE) path sampling method can readily sample multiple channels. In a first test, both the WE and transition path sampling methods are applied to two-dimensional model potentials. The comparison explains why the weighted ensemble approach will not be trapped in one channel. The WE approach is then used to sample the full transition path ensemble in implicitly solvated alanine dipeptide at two different temperatures. The ensembles are of sufficient quality to permit quantification of the fractional importance of each channel, even at $T = 300\text{K}$ when brute-force simulation is prohibitively expensive.

* Electronic mail: ddmz@pitt.edu

I. INTRODUCTION

Path sampling is an important strategy for studying study conformational transitions and activated processes. Some path sampling methods were applied to large biological system recently, for example, to study the conformational transitions of proteins. Conformational transition is important for protein to carry out their functions. But the pathway of these transitions are not easy to obtain either by experiments or simulations. Experiments, such as resolution studies by X-ray or NMR, are capable to determine the atomic structure of the stable states of protein, but not the transition states in the middle of the pathway. Conventional MD simulations are not good to study activated process because it involves a waiting step, which has been studied for a long time. The long waiting time to generate an ensemble of activated processes by straightforward brute-force approach is usually beyond the reach of today's computational power. In fact, if the activated process does take place, it is usually very quickly. Path sampling approaches take advantage of this property by focusing computer resources exclusively on rare transition events.

A number of path generating and path sampling approaches have been developed. Some *ad hoc* path generating methods have been applied for large biological systems, like the targeted molecular dynamics (TMD)^{1,2,3,4} and steered molecular dynamics (SMD).^{5,6,7} Another type of path sampling approaches focus on the optimal paths or the paths close to the optimal paths, such as the milestone method and the string method. Some path sampling methods take root in the idea of path integral methods, they sample the paths of rare events by evaluating their relative weights, such as, the transition path sampling (TPS),^{8,9,10} the dynamic importance sampling (DIMS),^{11,12,13} the transition interface sampling approach (TIS)^{14,15} and the random walk in path space by following a pseudo Langevin equation. In principle those methods based on the calculation of the weights of paths can generate properly distributed path ensembles. Not long ago Zuckerman, Jasnow and Zhang showed the Weighted Ensemble (WE) method^{16,17,18} is a very promising simulation approach to investigate conformational transitions, which can lead to correct path ensembles and reaction rate simultaneously. WE is also an exact algorithm for many dynamics types.¹⁹

Despite a number of successes, path sampling research still faces many challenges. One key difficulty is finding multiple reaction channels for a transition.²⁰ For example, the

transition path sampling approach, perhaps the best known method, can become trapped in a local minimum, and miss other channels.²⁰ The transition interface sampling method has the same drawback as its ancestor, TPS. Recently Bolhuis and van Erp tried to solve this problem by using the combination of the replica exchange and the transition interface sampling methods (RETIS).^{20,21}

The aim of this paper is to show the weighted ensemble method has no difficulty to sample multiple transition channels. The weighted ensemble method, which is based on unbiased replication and combination of brute-force simulations, does not have this drawback.

This paper is set up as follows. In Sec.II, we introduce the models and review the methods briefly. In Sec.III, first the weighted ensemble and transition path sampling methods are applied to two-dimensional model potentials. This simple example highlights the differences between the two methods sampling of multiple channels. Then the weighted ensemble approach is tested to find the transition events between two stable structures of alanine dipeptide at high temperature 500K. The path ensembles are compared with those gained from brute-force simulations. At last the weighted ensemble method samples the *full* transition ensemble of alanine dipeptide at room temperature 300K, which is prohibitively expensive for brute-force simulation. This paper ends with conclusions and discussion.

II. MODELS AND METHODS

A. Two-Dimensional Models

We will show the difference of how the WE approach and TPS approach sample multiple channels by using two-dimensional model potentials. The toy potentials U_1 and U_2 are inspired by Chen, Nash and Horing's work, and defined via

$$\begin{aligned}
 U(x, y)/k_B T = & \alpha(x^2 + y^2 - 1)^2 y^2 \\
 & - \exp\{-4[(x - 1)^2 + y^2]\} - \exp\{-4[(x + 1)^2 + y^2]\} \\
 & + \exp[8(x - 1.5)] + 4 \exp[-8(x + 1.5)] \\
 & + \exp[-4(y + 0.25)] + 16 \exp(-2x^2),
 \end{aligned} \tag{1}$$

with $\alpha = 20$ and $\alpha = 72$ respectively distinguishing U_1 and U_2 . The contours of these two potentials are shown in Fig. 1. They both have two channels connecting the left and right wells, which are shown by the red arrows. The only significant difference between these two potentials is the height of barrier separating two channels. Compared with the saddle points ($\sim 16k_B T$ and $17k_B T$ in Fig. 1), when $\alpha = 20$, the barrier height is about $3k_B T$. And when $\alpha = 72$, the barrier height is about $10k_B T$. The WE approach (using the horizontal position x as the progress coordinate) and TPS approach are applied to study the transition events of an over-damped Brownian particle from the left well to the right well. The initial state and the final state are defined as the regions where the potential satisfies $U(x, y) < (U_{min} + 2)$, where U_{min} is the lowest potential in the left and right wells.

B. Alanine Dipeptide

The second system on which we will test the weighted ensemble approach is alanine dipeptide (ace-ala-nme). The molecule is shown in Fig. 2. The principle variables describing the structure of alanine dipeptide are two backbone dihedral angles: Φ (C-N-C-C) and Ψ (N-C-C-N). Alanine dipeptide is frequently used for testing simulation methods and force fields.²² The reasons of choosing this molecule are as following. First it is one of the simplest molecules which contains two full peptide planes, so it will contains many structural features of protein backbones.²³ Second alanine dipeptide is small enough that its free energy surface can be studied thoroughly by different approaches.^{22,24,25,26,27,28,29,30} Third the conformational transitions of alanine dipeptide contains multiple channels and have been studied by several groups recently.^{4,13,31}

Several brute force simulations are performed first, using Langevin dynamics in the CHARMM program³² to find the energy minimum states. The simulations use the “united atom model” with the CHARMM parameter set 19 and implicit solvent ACE (analytical continuum electrostatics) model.³³ The dihedral angles of the four energy minimum states we find are shown in Table. I. Our locations of minima are somewhat different with the previous study,^{4,13,25,29} but it is known that the simulation of alanine dipeptide is very sensitive to solvent model.^{13,34} The WE approach is applied to study the transition events between state C_{7eq} and C_{7ax} . The initial state C_{7eq} is defined as the

area closed by the circle

$$[\Psi - (-77.9)]^2 + [\Phi - (138.4)]^2 = (40)^2, \quad (2)$$

and the final state C_{7ax} is the area closed the circle

$$[\Psi - (61.4)]^2 + [\Phi - (-71.4)]^2 = (20)^2, \quad (3)$$

as shown in Fig. 3.

We choose the dihedral angles Ψ and Φ as the progress coordinates for the WE approach. The two-dimensional space of dihedral angles is cut into a 12×12 grid, with 20 simulations allowed in each grid. After every $\tau = 100\text{fs}$, the embedded CHARMM simulations are paused, and the simulations are combined and split without bias. The weighted ensemble program was stopped after 2500τ .

III. RESULTS

A. Two-Dimensional Models

The TPS approach and the WE approach sample the paths in multiple channels in different ways. To show the difference, we use the y position of path crossing the $x = 0$ section last time to identify which channel it belongs to, and plot this position for each path versus the path index (simulation time). Fig. 4(a) and Fig. 4(b) show the path switching of TPS method for potentials U_1 and U_2 . The TPS method, as a Monte Carlo simulation, samples the channels (local minimum in the path space) one by one. For potential U_1 , the switch happens about every 10^4 paths, For potential U_2 , because of the higher barrier, the switch happens approximately every $5 \cdot 10^6$ paths, which is much less frequent. This reveals how, for higher barriers, the sampling can get “trapped” and could be biased. Fig. 4(c) and Fig. 4(d) show the path switching of WE method for these two potentials. The frequent switches suggest that the Weighted Ensemble method, after a short transient period, generates different types of paths simultaneously in both potentials.

Both methods are used to get the distribution of transition-event durations. The definition of the transition-event duration is the time interval between the last time the

Brownian particle leave the initial state and the first time it reaches the final state. It is the time the Brownian particle uses to finish the transition. The WE simulations are stopped after they generate 10^6 paths. The TPS simulations are stopped after $2.5 \cdot 10^7$ paths are obtained. The results are shown in Fig. 5. For small barrier (U_1), both methods get the correct distribution by using 10^6 paths. But for potential U_2 , because of the high barrier separating the two types of paths, a “short” transition path sampling simulation yields an incorrect distribution after 10^6 paths have been sampled. In that case, the transition path sampling is trapped by the local minimum in path space; see Fig. 4(b).

B. Alanine Dipeptide

The WE method is applied to study the transition events between state C_{7eq} and C_{7ax} of alanine dipeptide under temperature 300K and 500K. The brute-force simulations are also used to get the transition paths under 500K, and the results are compared with those given by WE method.

1. The Transition Rate

The reaction rate k is an important quantity for chemical and biological reactions and transitions. If the first passage time is too long, this quantity will be impossible to obtain by brute-force simulations. The WE method can yield the path ensemble and reaction rate simultaneously. Under 500K brute-force simulations obtained the reaction rate $k_{BF} = 1.5 \times 10^{-1}/ns$. And WE method obtained the reaction rate $k_{WE} = 1.4 \times 10^{-1}/ns$. They are in good agreement. When temperature is 300K, the reaction rate obtained by WE method is $k_{WE} = 1.6 \times 10^{-3}/ns$, which means compared with 500K, the transition is about 100 times more difficult to happen now.

2. Paths in Different Channels

The path ensemble will be studied in an enlarged dihedral plane (Fig. 6) which clearly shows alternative and continuous paths. The transition paths can be roughly divided into four types according to which of the (physically equivalent) C_{7ax} states they end in this

extended dihedral plane. These four types of paths correspond to the combinations of clockwise and anticlockwise rotational directions of dihedral angles Ψ and Φ . Different types of paths pass different barriers; therefore they belong to different channels. Under 300K, approximately 100 paths randomly chosen (based on their weights) are plotted in Fig. 6.

The distributions of these four types of paths are listed in Table II. Once again under 500K, the results given by WE method and brute-force are in good agreements. Because the enlarged plane is infinite, there are also infinite final C_{7ax} states in it. Some paths ended out of these four closest final states, the distributions of them are listed in the "others" column in the table. The path ensemble connecting to the "lower right" C_{7ax} state is the most important type. And when the temperature decreased, it became more dominant. From Fig. 6, we can tell there are two channels in this type of paths. They are divided apart by the high barrier around $\Psi = 0$ and $\Phi = 0$.

IV. CONCLUSIONS

Our results show that the weighted ensemble (WE) path sampling approach is naturally suited to the fundamental problem of sampling multiple reaction channels in molecular systems. The ability to sample multiple channels is more than an abstract statistical mechanics issue: after all, it would not seem possible to find an "optimal" path without the ability to fully traverse path space.

Our studies employed toy systems to illustrate the basic mechanisms underlying WE and the more familiar transition path sampling (TPS) methods, and then focused on a molecular system in detail. In the toy systems, the differences between WE and TPS were clear: because TPS is a Monte Carlo simulation, it can be trapped in one path channel. By contrast, the WE method is based on the unbiased replication and combination and of brute-force simulations, and cannot be trapped. The WE approach was also employed to find transition events between the C_{7eq} and C_{7ax} states of atomistic alanine dipeptide at temperature 500K. The results were checked and supported by brute-force simulations. Finally, the WE method was employed to study the same transition at room temperature 300K. A high-quality transition path ensemble including multiple, clearly separated channels was obtained.

We do not know of any previous report of a statistically rigorous path ensemble of alanine dipeptide. Put another way, this appears to be the first report of the fractional importance of the various channels to the path ensemble, despite the small size of the molecule.

In the long term, we believe WE path sampling has key strengths, and also some weaknesses. WE is easy to implement and suitable for use with almost type of stochastic dynamics — including MD with a stochastic thermostat.¹⁹ The multiple-trajectory “architecture” of WE makes it straightforward to parallelize. WE does not require precise knowledge of a reaction coordinate^{18,19} — indeed, no “targeting” was used in the alanine dipeptide simulations reported here. Further, WE sampling simultaneously provides the reaction rate and path ensemble. Like TPS, however, WE produces correlated trajectories in the path ensembles generated — although this does not prevent the method from achieving efficiency.¹⁸ Ultimately, the great flexibility of the approach, especially with regard to binning strategies^{18,19} suggests we are only at the beginning of appreciating the possibilities of Huber and Kim’s seminal strategy.¹⁶

-
- ¹ J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, *Mol. Simul.* **10**, 291 (1993).
- ² J. Schlitter, M. Engels, and P. Krüger, *J. Mol. Graph.* **12**, 84 (1994).
- ³ A. van der Vaart and M. Karplus, *J. Chem. Phys.* **122**, 114903 (2005).
- ⁴ A. van der Vaart and M. Karplus, *J. Chem. Phys.* **126**, 164106 (2007).
- ⁵ S. Izrailev, S. Stepaniants, M. Balsara, Y. Oono, and K. Schulten, *Biophys. J.* **72**, 1568 (1997).
- ⁶ H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, *Biophys. J.* **75**, 662 (1998).
- ⁷ P. E. Marszalek *et al.*, *Nature* **402**, 100 (1999).
- ⁸ C. Dellago, P. G. Bolhuis, and D. Chandler, *J. Chem. Phys.* **108**, 9236 (1998).
- ⁹ C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- ¹⁰ P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- ¹¹ D. M. Zuckerman and T. B. Woolf, *the Journal of Chemical Physics* **111**, 9475 (1999).
- ¹² D. M. Zuckerman and T. B. Woolf, *Phys. Rev. E* **63**, 016702 (2000).
- ¹³ H. Jang and T. B. Woolf, *J. Comput. Chem.* **27**, 1136 (2006).
- ¹⁴ T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **118**, 7762 (2003).
- ¹⁵ T. S. van Erp and P. G. Bolhuis, *J. Comput. Phys.* **205**, 157 (2005).
- ¹⁶ G. A. Huber and S. Kim, *Biophys. J.* **70**, 97 (1996).
- ¹⁷ A. Rojnuckarin, S. Kim, and S. Subramaniam, *Proc. Natl. Acad. Sci. USA* **95**, 4288 (1998).
- ¹⁸ B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *Proc. Natl. Acad. Sci. USA* **104**, 18043 (2007).
- ¹⁹ D. M. Zuckerman and D. Jasnow, (2008), <http://www.citebase.org/abstract?id=oai:arXiv.org:0810.1963>.
- ²⁰ P. G. Bolhuis, *J. Chem. Phys.* **129**, 114108 (2008).
- ²¹ T. S. van Erp, *Comput. Phys. Commun.* **179**, 34 (2008).
- ²² D. J. Tobias and C. L. Brooks, *J. Phys. Chem.* **96**, 3864 (1992).
- ²³ C. Branden and J. Tooze, *Introduction to Protein Structure*, 1st ed. (Garland Pub., 1991).
- ²⁴ D. M. Philipp and R. A. Friesner, *J. Comput. Chem.* **20**, 1468 (1999).
- ²⁵ J. Apostolakis, P. Ferrara, and A. Caflisch, *J. Chem. Phys.* **110**, 2099 (1999).
- ²⁶ T. Marrone, M. Gilson, and J. McCammon, *J. Phys. Chem.* **100**, 1439 (1996).
- ²⁷ A. Drozdov, A. Grossfield, and R. Pappu, *J. Am. Chem. Soc.* **126**, 2574 (2004).
- ²⁸ P. E. Smith, B. M. Pettitt, and M. Karplus, *J. Phys. Chem.* **97**, 6907 (1993).
- ²⁹ P. E. Smith, *J. Chem. Phys.* **111**, 5568 (1999).

- ³⁰ R. Vargas, J. Garza, B. Hay, and D. Dixon, J. Phys. Chem. A **106**, 3213 (2002).
- ³¹ D. Branduardi, F. L. Gervasio, and M. Parrinello, J. Chem. Phys. **126**, 054103 (2007).
- ³² B. R. Brooks *et al.*, J. Comput. Chem. **4**, 187 (1983).
- ³³ M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
- ³⁴ T. B. Woolf, Chem. Phys. Lett. **289**, 433 (1998).

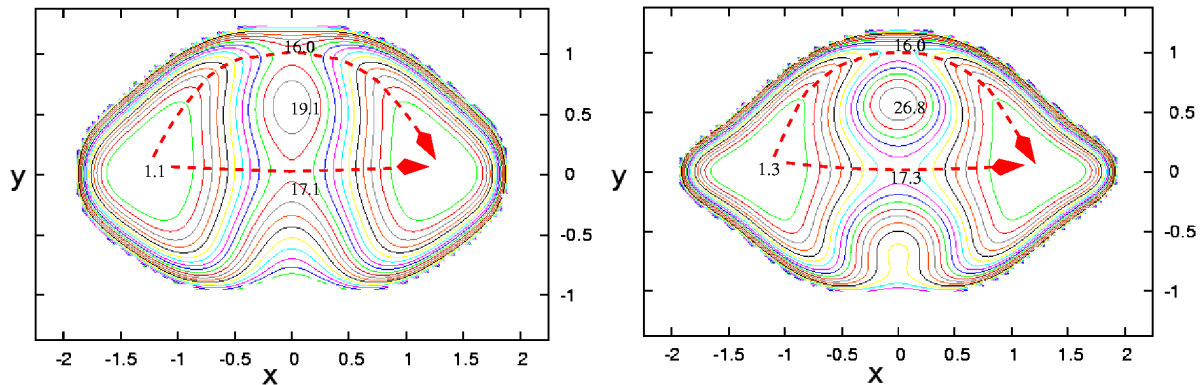


Figure 1: Contours of two-dimensional toy potentials. The left panel is the potential U_1 with $\alpha = 20$, the right panel is the potential U_2 with $\alpha = 72$. The red arrows show the different channels connecting the left and right wells. The numbers among the contours indicate energy values in $k_B T$ units for extrema and saddles.

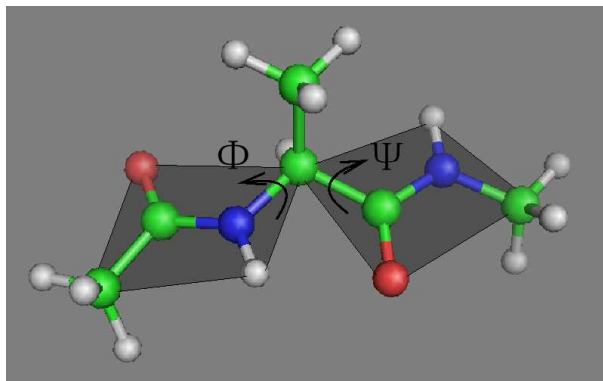


Figure 2: Alanine dipeptide molecule ($\text{CH}_3\text{CO} - \text{Ala} - \text{NHCH}_3$). Different atoms are shown by different colors. White: Hydrogen; Red: Oxygen; Green: Carbon; Blue: Nitrogen.

α_L	C_{7eq}	α_R	C_{7ax}
(55.1,46.4)	(-77.9,138.4)	(-75.6,-39.9)	(61.4,-71.4)

Table I: Four energy minima of alanine dipeptide.

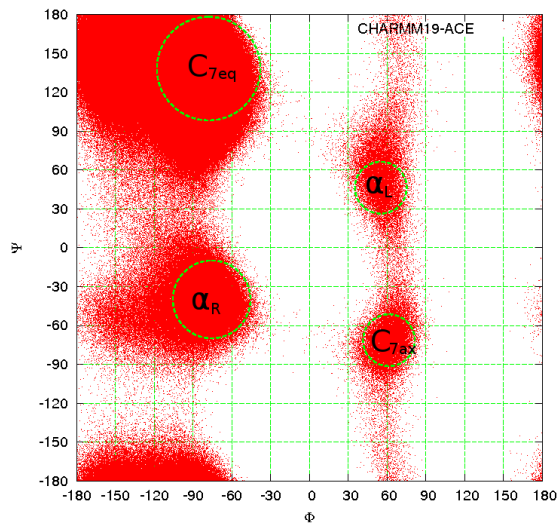


Figure 3: Four stable states of alanine dipeptide. The dihedral angles Ψ and Φ are chosen as the progress coordinates for the WE approach. The two-dimensional space of dihedral angles is divided into a 12×12 grid for WE simulation, with 20 simulations allowed in each grid.

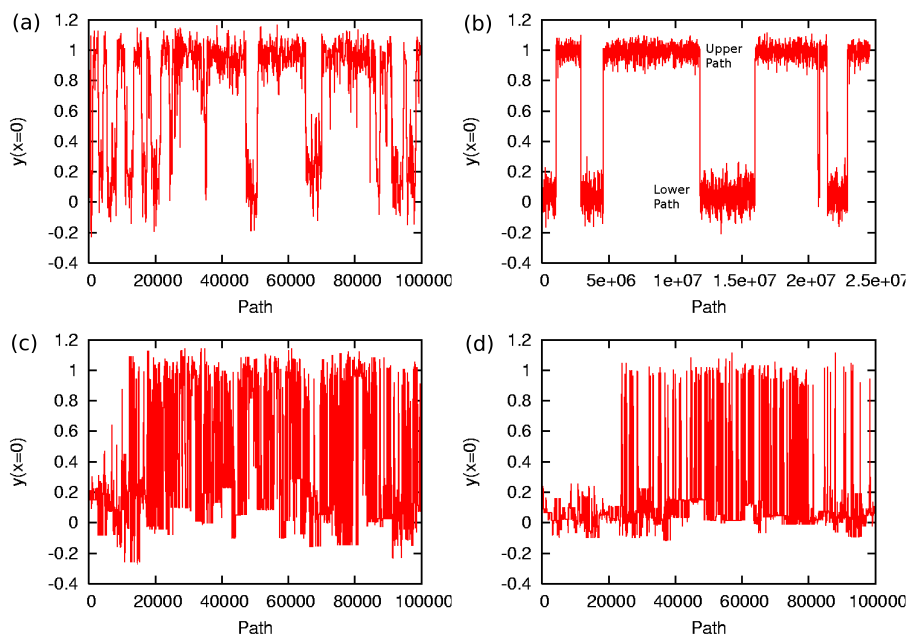


Figure 4: Channel-switching in TPS and WE simulations of toy models. Both methods were applied to the two toy potentials, U_1 and U_2 of Fig. 1. The various panels show (a)TPS for U_1 , (b)TPS for U_2 , (c)WE for U_1 and (d)WE for U_2 . The TPS method samples the channels one by one. In panel (a), for potential U_1 , the switch happens about every 10^4 paths. In panel (b), for potential U_2 , the switch happens approximately every $5 \cdot 10^6$ paths. But in panel (c) and (d), after a short transient period, WE simulations generate different types of paths simultaneously. Note the greatly enlarged horizontal scale in panel (b).

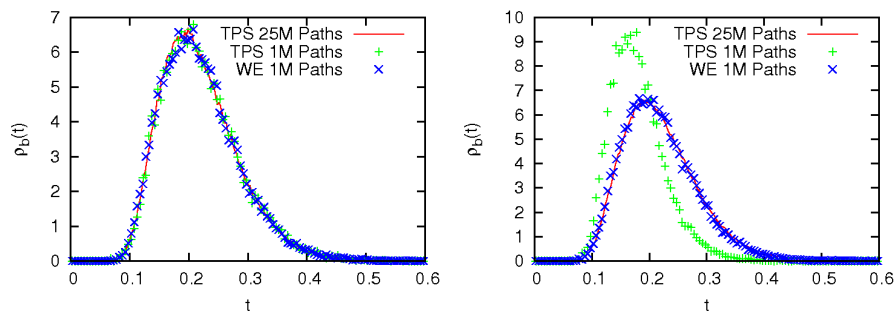


Figure 5: Distribution of transition-event durations in toy models. The left panel is for potential U_1 (smaller barrier), the right panel is for potential U_2 (larger barrier). Notice in the right panel, a “short” transition path sampling simulation yields an incorrect distribution.

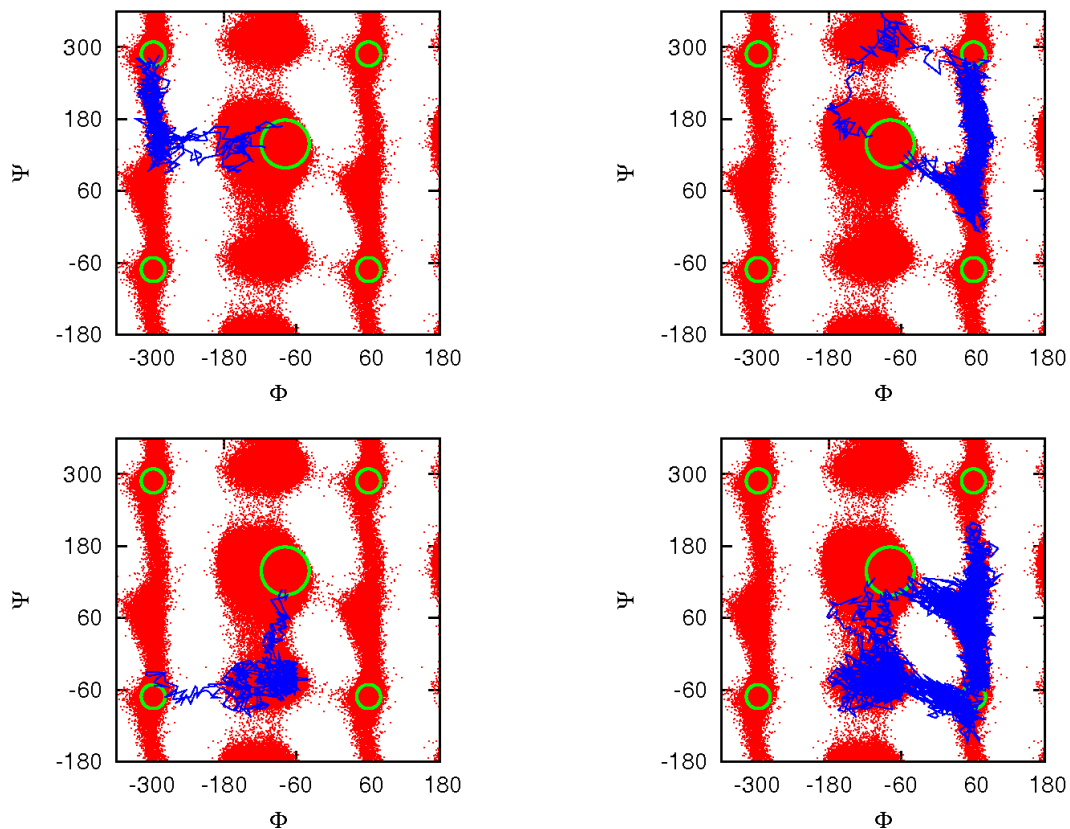


Figure 6: Four types of transition paths between the states C_{7eq} and C_{7ax} of alanine dipeptide at 300K. Approximately 100 paths are randomly chosen based on their weights (resampled) from the full path ensemble. Different paths traverse different barriers, and therefore belong to different channels.

	upper left	upper right	lower left	lower right	others
BF 500K	4.2%	25.6%	5.3%	64.1%	0.8%
WE 500K	6.0%	21.4%	4.9%	67.4%	0.4%
WE 300K	0.6%	12.3%	0.9%	86.2%	0.0%

Table II: Distributions of the four principal channels in alanine dipeptide. The position of the final $C_{7_{ax}}$ state in the extended plane of Fig. 6 was used to identify the channel. The fraction of paths which did not end in one of these four closest final states are listed in the “others” column.